

# **APLIKASI DATAMINING UNTUK MEMPREDIKSI TINGKAT KELULUSAN MAHASISWA MENGGUNAKAN ALGORITMA *APRIORI* DI IBI DARMAJAYA BANDAR LAMPUNG**

**Nurjoko<sup>1</sup>      Hendra Kurniawan<sup>2</sup>**

<sup>1,2</sup>Jurusan Sistem Informasi  
<sup>1,2</sup>Informatics and Business Institute Darmajaya  
E-mail: <sup>1</sup>[nurjoko.mti@gmail.com](mailto:nurjoko.mti@gmail.com)  
E-mail: <sup>2</sup>[hendra.kurniawan@darmajaya.ac.id](mailto:hendra.kurniawan@darmajaya.ac.id)

## **ABSTRAK**

Pertumbuhan yang pesat dari akumulasi data telah menciptakan kondisi kaya akan data tapi minim informasi. Data mining merupakan penambangan atau penemuan informasi baru dengan mencari pola atau aturan tertentu dari sejumlah data dalam jumlah besar yang diharapkan dapat mengatasi kondisi tersebut. Dengan memanfaatkan data induk mahasiswa dan data kelulusan mahasiswa, diharapkan dapat menghasilkan informasi tentang tingkat kelulusan dengan data induk mahasiswa melalui teknik data mining. Kategori tingkat kelulusan diukur dari lama studi dan IPK. Algoritma yang digunakan adalah algoritma apriori, informasi yang ditampilkan berupa nilai *support* dan *confidence* dari masing-masing kategori tingkat kelulusan.

**Kata Kunci:** data mining, algoritma apriori, tingkat kelulusan, support, confidence

## **ABSTRACT**

*The rapid growth of data accumulation has created conditions data-rich but minimal information. Data mining is a mining or the discovery of new information by looking for patterns or particular rules of a number of large amounts of data are expected to tackle the condition. By leveraging master data and data graduation student, is expected to yield information about the graduation rate of students with master data through data mining techniques. Categories graduation rate is measured from the time of study and GPA. The algorithm used is the a priori algorithm, the value of the information displayed in the form of support and confidence from each category graduation rates.*

**Keywords:** data mining, algorithms priori, graduation rates, support, confidence

## **1. PENDAHULUAN**

### **1.1 Latar Belakang Masalah**

Perguruan tinggi saat ini dituntut untuk memiliki keunggulan bersaing dengan memanfaatkan semua sumber daya yang dimiliki. Selain sumber daya sarana, prasarana, dan manusia, sistem informasi adalah salah satu sumber daya yang dapat digunakan untuk meningkatkan keunggulan bersaing. Sistem informasi dapat digunakan untuk mendapatkan, mengolah dan menyebarkan informasi untuk menunjang kegiatan operasional sehari-hari sekaligus menunjang kegiatan pengambilan keputusan strategis. Di dalam peraturan akademik IBI Darmajaya bidang pendidikan disebutkan bahwa “Program

Sarjana (S1) reguler adalah program pendidikan akademik setelah pendidikan menengah, yang memiliki beban studi sekurang-kurangnya 144 (seratus empat puluh empat) sks (satuan kredit semester) dan sebanyak-banyaknya 160 (seratus enam puluh ) sks yang dijadwalkan untuk 8 (delapan) semester dan dapat ditempuh dalam waktu kurang dari 8 (delapan) semester dan paling lama 14 (empat belas) semester” (Peraturan Akademik,). Berdasarkan data peserta wisuda Program Sarjana (S1) di IBI Darmajaya banyak alumni yang menempuh masa studi lebih dari 8 semester. Hal ini menunjukkan bahwa masih banyak mahasiswa Program Sarjana (S1) yang menempuh lama studi lebih dari yang dijadwalkan yaitu 8 semester. Oleh karena itu, dengan memanfaatkan data induk mahasiswa dan data kelulusan mahasiswa, dapat diketahui informasi tingkat kelulusan mahasiswa melalui teknik *data mining*.

## 1.2 Pendekatan Teori

*Data mining* membantu perusahaan untuk mendapatkan pola dari data-data yang tersimpan di dalam basis data perusahaan. Pengetahuan yang diperoleh tersebut akan menjadi pedoman dalam mengambil tindakan-tindakan bisnis sebagai upaya pemeliharaan dan peningkatan tingkat kompetitif bisnis perusahaan. Walaupun sudah banyak perangkat lunak yang menawarkan kemampuan dalam proses *data mining*, keterlibatan manusia sangat dibutuhkan dalam setiap fase proses *data mining* itu sendiri. Pemahaman terhadap model statistik dan matematik yang digunakan dalam perangkat lunak sangat dituntut.

Banyak pihak yang telah mendefinisikan *data mining*. Berikut beberapa definisi *data mining*:

“*Data mining* merupakan suatu proses pencarian pola dari data-data dengan jumlah yang sangat banyak yang tersimpan dalam suatu tempat penyimpanan dengan menggunakan teknologi pengenalan pola, teknik statistik, dan matematik.” (Gartner Group)

“*Data mining* merupakan analisis dari sekumpulan data yang diamati (sangat besar) untuk menemukan hubungan yang tidak terduga dan merangkum data dengan cara yang baru yang dapat dipahami dan berguna bagi pemilik data.” (Hand, et al) Fungsi *data mining* dapat menjalankan fungsi-fungsi berikut:

### Deskripsi

Pola dan trend data sering dideskripsikan. Deskripsi tersebut sangat membantu dalam menjelaskan pola dan trend yang terjadi. Model *data mining* harus setransparan mungkin, dimana hasilnya dapat mendeskripsikan pola dengan jelas.

### Estimasi

Estimasi mirip dengan klasifikasi kecuali variabel target-nya numerik ketimbang kategorikal. Model yang dibangun menggunakan *record* yang lengkap, yang menyediakan nilai variabel target dan *predictor*. Untuk observasi yang baru, estimasi nilai variabel target ditentukan, berdasarkan nilai-nilai *predictor*.

### Prediksi

Prediksi mirip dengan klasifikasi dan estimasi, kecuali dalam prediksi, hasil terjadi di masa datang.

### **Klasifikasi**

Dalam klasifikasi, variabel target-nya merupakan kategorikal. Model *data mining* memeriksa set *record* yang besar, tiap *record* mempunyai informasi variabel target dan set input atau variabel *predictor*.

### **Clustering**

*Clustering* merupakan pengelompokkan *record*, observasi, atau kasus ke dalam kelas-kelas objek yang mirip. *Clustering* berbeda dengan klasifikasi dimana dalam *clustering* tidak terdapat variabel target. *Clustering* mencoba menyegmentasi seluruh set data ke dalam *subgroup* atau *cluster* yang relatif homogen, dimana kemiripan antar *record* dalam *cluster* dimaksimalkan dan kemiripan *record* di luar *cluster* diminimasi.

### **Asosiasi**

Asosiasi merupakan suatu tugas untuk menemukan atribut-atribut yang “terjadi” bersamaan. Tugas asosiasi mencoba untuk menemukan aturan untuk mengkuantifikasi hubungan antara dua atau lebih atribut. Aturan asosiasi berbentuk “If *antecedent*, then *consequent*”, bersama-sama dengan ukuran *support* dan *confidence* yang berhubungan dengan aturan.

## **2. METODOLOGI PENELITIAN**

### **2.1 METODE DATA MINING ASSOCIATION RULES**

#### **ANALISIS DATA MINING**

Dalam penelitian ini akan dicari nilai *support* dan *confidence* dari hubungan tingkat kelulusan dengan data induk mahasiswa. Tidak semua data induk siswa akan dicari hubungannya dengan data kelulusan, hanya beberapa atribut yang kira-kira berguna dan sebarannya tidak terlalu acak. Karena data yang terlalu acak akan membuat proses *mining* memakan waktu lama dan tingkat hubungannya pun rendah. Data induk mahasiswa yang akan dicari hubungannya meliputi asal sekolah, kota asal sekolah, dan program studi. Adapun yang akan diproses *mining* meliputi:

#### **1. Hubungan tingkat kelulusan dengan asal sekolah**

Dari atribut proses masuk dan asal sekolah dicari hubungan tingkat kelulusan dengan asal sekolah yang melalui proses masuk PMB dengan harapan dapat mengetahui tingkat keberhasilan mahasiswa dengan sekolah tertentu.

#### **2. Hubungan tingkat kelulusan dengan kota asal sekolah**

Hubungan tingkat kelulusan dengan asal kota bermanfaat untuk mengetahui daerah-daerah mana yang mempunyai tingkat keberhasilan tinggi ataupun rendah. Diasumsikan bahwa kota asal sekolah merupakan kota tempat asal mahasiswa.

#### **3. Hubungan tingkat kelulusan dengan program studi**

Dari atribut program studi dapat diketahui hubungan tingkat kelulusan dan program studi untuk mengetahui tingkat kelulusan program studi.

## SUMBER DATA

Data yang digunakan dalam penelitian ini terdiri dari dua sumber data, yaitu data Induk Mahasiswa dan data Kelulusan.

### 1. Data Induk Mahasiswa

Data induk mahasiswa adalah data mahasiswa yang didata ketika mahasiswa pertama kali masuk perguruan tinggi setelah melakukan registrasi ulang. Data yang dicatat adalah identitas pribadi mahasiswa dan identitas sekolah asal mahasiswa.

### 2. Data Kelulusan

Data Kelulusan adalah data mahasiswa yang telah dinyatakan lulus. Data yang dicatat adalah identitas mahasiswa dan data kelengkapan kelulusan. Data induk mahasiswa yang diambil dalam sampel adalah data mahasiswa angkatan 2004, 2005, 2006, 2007 dan 2008. Hal ini didasarkan pada kebutuhan data yang akan di hubungkan dengan data kelulusan, dengan asumsi bahwa mahasiswa angkatan 2004 - 2007 akan lulus dari rentang waktu tahun 2008-2011. Sedangkan data kelulusan yang diambil adalah data kelulusan dari tahun 2008 sampai 2011. Kedua data tersebut diperoleh dari bagian akademik IBI Darmajaya. Data yang diambil hanya dari mahasiswa sarjana (S1).

## DATA YANG DIGUNAKAN

Dalam penelitian ini dicari hubungan beberapa atribut dari data induk mahasiswa dengan tingkat kelulusan. Karena tidak semua tabel digunakan maka perlu dilakukan pembersihan data agar data yang akan diolah benar-benar relevan dengan yang dibutuhkan. Pembersihan ini penting guna meningkatkan performa dalam proses *mining*. Cara pembersihan dengan menghapus atribut yang tidak terpakai dan menghapus data-data yang tidak lengkap isian. atribut yang digunakan terdiri dari atribut pada data kelulusan dan pada data induk mahasiswa. Atribut yang digunakan dalam data induk mahasiswa meliputi:

1. Atribut NIM digunakan sebagai *primary key* untuk menghubungkan dengan data kelulusan.
2. Atribut nama asal sekolah digunakan untuk proses *mining* guna mengetahui hubungan antara tingkat kelulusan dengan asal sekolah.
3. Atribut kota asal sekolah digunakan untuk proses *mining* guna mengetahui hubungan tingkat kelulusan dengan kota asal mahasiswa.

Atribut yang digunakan dalam data kelulusan meliputi:

1. NIM digunakan sebagai *primary key* untuk menghubungkan dengan data induk mahasiswa.
2. Indeks Prestasi Kumulatif (IPK) digunakan sebagai ukuran tingkat kelulusan mahasiswa
3. Lama studi digunakan sebagai ukuran tingkat kelulusan mahasiswa.

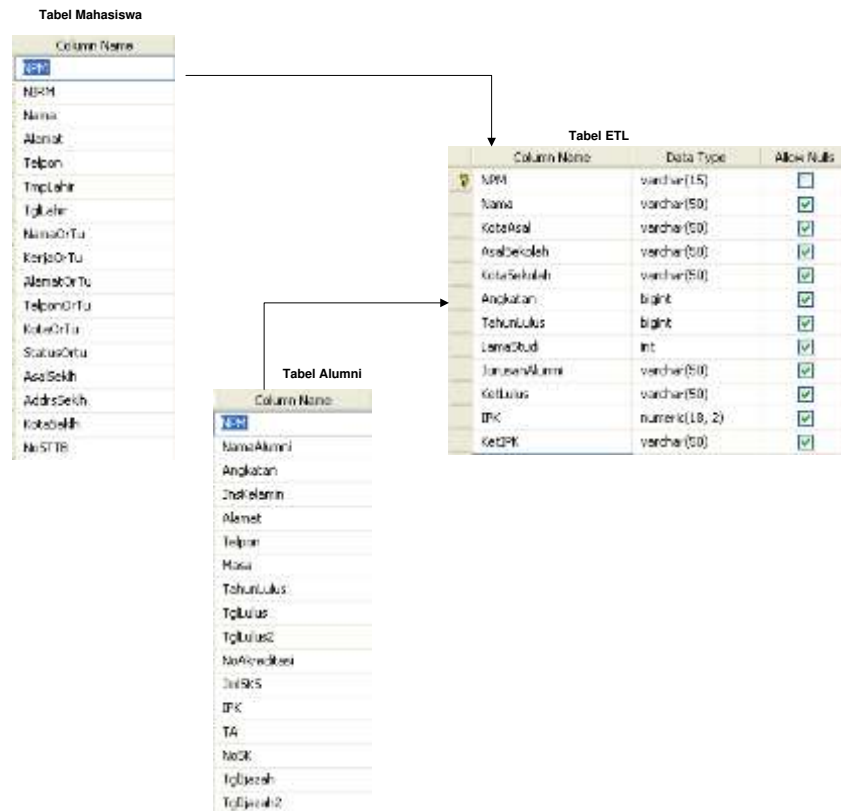
Berikut proses-proses yang terdapat pada Aplikasi *Data Mining*:

### 1. Import Data

Proses *import* data adalah proses *load* data dari *database* kelulusan dan *database* induk mahasiswa ke *data warehouse*. Semua data akan dimasukkan tanpa ada penyaringan.

## 2. *Cleaning, Integrasi, Selection, dan transformasi*

- a) *Cleaning* data merupakan proses menghilangkan *noise* dan data yang tidak konsisten atau data yang tidak relevan. Pada umumnya data yang diperoleh, baik dari *database* suatu perusahaan maupun hasil eksperimen, memiliki isian-isian yang tidak sempurna seperti data yang hilang, data yang tidak valid atau juga hanya sekedar salah ketik. Dalam tahap ini semua data yang akan digunakan baik data kelulusan, data induk mahasiswa maupun data nilai semester dibersihkan dari *record* data yang tidak mempunyai atribut lengkap. Selain pembersihan *record* data yang tidak *valid*, juga dilakukan penghapusan atribut yang tidak dipakai, misalnya atribut gaji orang tua, nama orang tua dan lain-lain. Pembersihan data juga akan mempengaruhi performansi dari sistem *data mining* karena data yang ditangani akan berkurang jumlah dan kompleksitasnya.
  - b) *Integrasi* data merupakan penggabungan data dari berbagai *database* ke dalam satu *database* baru. Data induk mahasiswa, data nilai dan data kelulusan tidak disimpan dalam satu *database*, *integrasi* data dilakukan pada atribut-atribut yang mengidentifikasi entitas-entitas dengan satu atribut unik yaitu NIM.
  - c) *Selection data* adalah proses menyeleksi atribut apa yang akan diproses pada mining selanjutnya.
  - d) *Transformasi* data merupakan proses mengubah data atau digabung ke dalam format yang sesuai untuk diproses dalam *data mining*. Beberapa metode *data mining* membutuhkan format data yang khusus sebelum bisa diaplikasikan. Dalam Aplikasi *Data Mining* ini, data yang dirubah yaitu lama studi dan IPK untuk mengukur tingkat kelulusan. Atribut lama studi dan IPK dibagi menjadi beberapa interval.
3. **Proses *mining***, proses masuk merupakan proses *mining* untuk mengetahui hubungan tingkat kelulusan dengan proses masuk mahasiswa.
4. **Proses *mining* asal kota**, merupakan proses *mining* untuk mengetahui hubungan tingkat kelulusan dengan asal kota mahasiswa, digunakan data kota asal sekolah dengan asumsi kota asal sekolah merupakan kota asal mahasiswa.
5. **Proses *mining* program studi**, merupakan proses *mining* untuk mengetahui hubungan tingkat kelulusan dengan program studi. Program studi digunakan untuk proses *mining* guna mengetahui hubungan tingkat kelulusan dengan program studi.



Gambar 1. Struktur Tabel Yang Digunakan

Atribut yang digunakan dalam table ETL dapat dijelaskan sebagai berikut:

1. NPM  
Atribut ini digunakan sebagai *primary key* untuk menghubungkan antara table induk mahasiswa dan table alumni. Atribut ini diambil dari table AlumniMhs, karena data alumni yang dijadikan sebagai acuan dalam proses mining.
2. Nama  
Atribut ini digunakan untuk mengambil nama alumni. Atribut ini diambil dari table AlumniMhs.
3. Kota Asal  
Atribut kota asal digunakan untuk mengetahui kota asal alumni. Atribut ini diambil dari table Induk mahasiswa mengingat informasi ini diambil pada saat mahasiswa mendaftar.
4. Asal Sekolah  
Atribut ini digunakan untuk mengetahui asal sekolah. Atribut ini diperoleh dari table induk mahasiswa.
5. Kota Sekolah  
Atribut ini digunakan untuk mengetahui kota sekolah asal mahasiswa. Atribut ini diambil dari table induk mahasiswa.
6. Angkatan  
Angkatan digunakan untuk mengetahui tahun kapan mahasiswa tersebut masuk ke perguruan tinggi. Atribut ini diambil dari table AlumniMhs.

7. Tahun Lulus  
Atribut ini digunakan untuk mengathui tahun kapan mahasiswa tersebut lulus. Atribut ini diambil dari table AlumniMhs.
8. Lama Studi  
Atribut lama studi digunakan untuk menyimpan hasil transformasi lama studi mahasiswa.
9. Jurusan Alumni  
Atribut ini digunakan untuk mengetahui jurusan alumni. Atribut ini diambil dari table AlumniMhs.
10. Keterangan Lulus  
Atribut ini digunakan untuk menyimpan data hasil transformasi lama studi.
11. IPK  
Atribut ini digunakan untuk mengetahui ipk mahasiswa pada saat lulus. Atribut ini diambil dari table AlumniMhs.
12. Keterangan IPK  
Atribut ini digunakan untuk menyimpan hasil dari transformasi IPK.

### PROSES INTEGRASI DATA

Tabel yang digunakan dalam penelitian ini berada dalam 1 server sehingga memudahkan dalam proses *Extract Transform Load* (ETL).

### PROSES TRANSFORMASI DATA

Transformasi data merupakan proses pengubahan atau penggabungan data ke dalam format yang sesuai untuk diproses dalam data mining. Seringkali data yang akan digunakan dalam proses data mining mempunyai format yang belum langsung bisa digunakan, oleh karena itu perlu dirubah formatnya.

1. Transformasi IPK  
Transformasi ini digunakan untuk menentukan predikat IPK, dengan kategori seperti pada tabel 1.

Tabel 1. Transformasi IPK

Indeks Prestasi Kumulatif	Predikat
2,00 – 2,75	Memuaskan
2,76 – 3,50	Sangat memuaskan
3,51 – 4,00	Dengan pujian (cumlaude)

2. Transformasi Keterangan Lulus  
Transformasi ini digunakan untuk menentukan tingkat kelulusan, dengan kategori seperti pada tabel 2.

Tabel 2. Transformasi Tingkat Kelulusan

Lama Studi	Jenjang	Tingkat Kelulusan
< 4	S1	Melebihi Batas Studi

=4	S1	Sesuai Batas Studi
> 4	S1	Cepat
< 3	D3	Melebihi Batas Studi
= 3	D3	Sesuai Batas Studi
> 3	D3	Cepat

## PROSES MINING

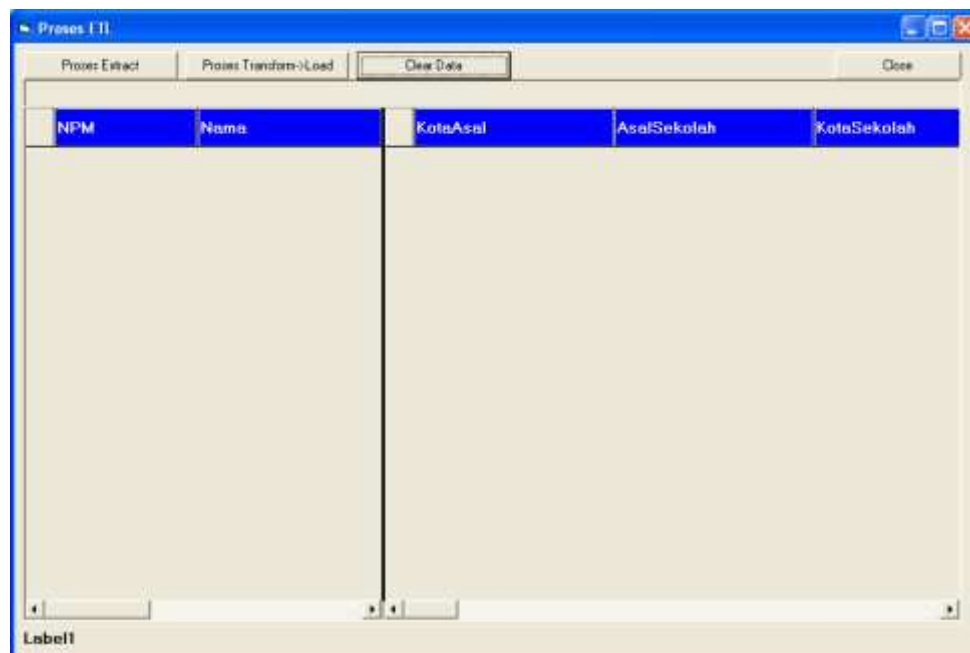
Untuk proses *mining* dilakukan dengan menggunakan MS. SQL Server 2008 Data Mining. Dengan menggunakan *tools* ini proses mining dapat dengan mudah diproses dan dihasilkan output yang mudah untuk dibaca.

## 3. HASIL DAN PEMBAHASAN

### 3.1 HASIL

Tahapan dalam penelitian ini dimulai dengan proses *Extract, Transform, Load* (ETL) dengan mengambil data dari tabel induk mahasiswa dan tabel alumni yang kemudian dimasukkan ke dalam tabel\_ETL. Setelah beberapa data penting berupa kolom-kolom tertentu didefinisikan dan tabel\_ETL telah berisi data, maka langkah selanjutnya adalah melakukan transformasi dengan mengubah beberapa atribut dalam tabel\_ETL menjadi data yang sesuai dengan rancangan transformasi.

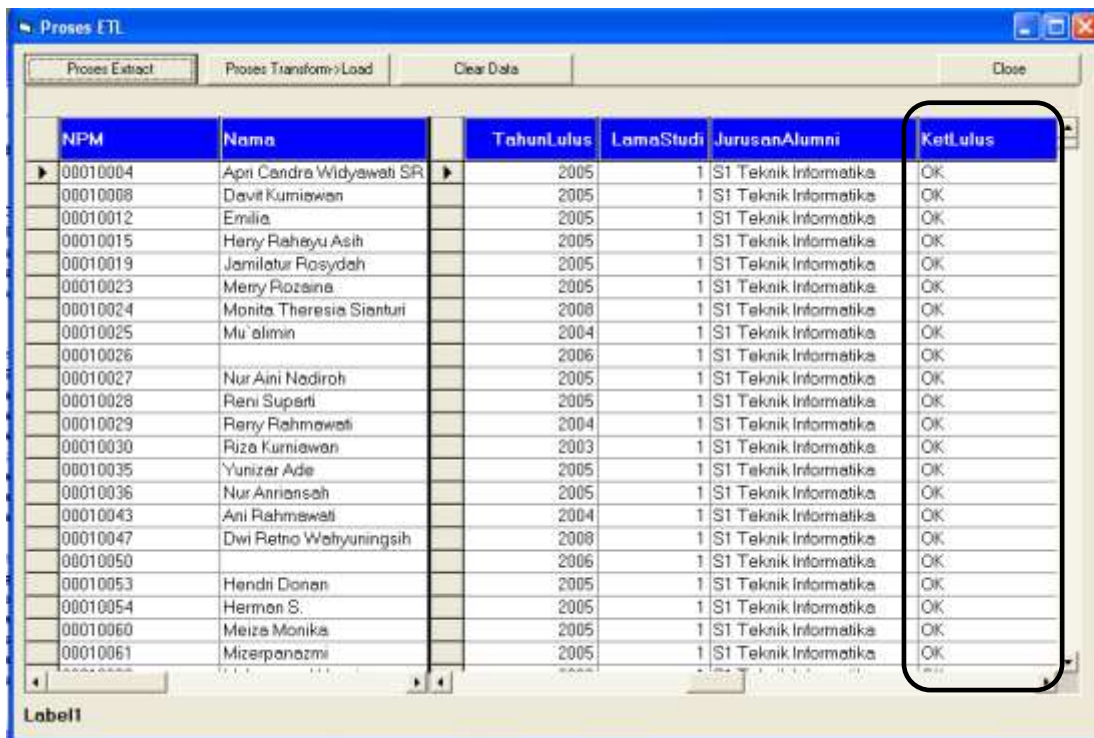
Proses *extract* data dilakukan melalui bantuan program yang dibangun menggunakan MS. *Visual Basic 6* seperti dapat dilihat pada gambar 2.



Gambar 2. Proses *Extract*

Untuk melakukan proses *extract*, cukup dilakukan dengan cara klik tombol Proses Extract. Setelah tahap ini dilakukan, maka akan dimunculkan sejumlah data yang diperoleh dari proses *extract*. Hasil dari proses *extract* dapat dilihat pada gambar 3.

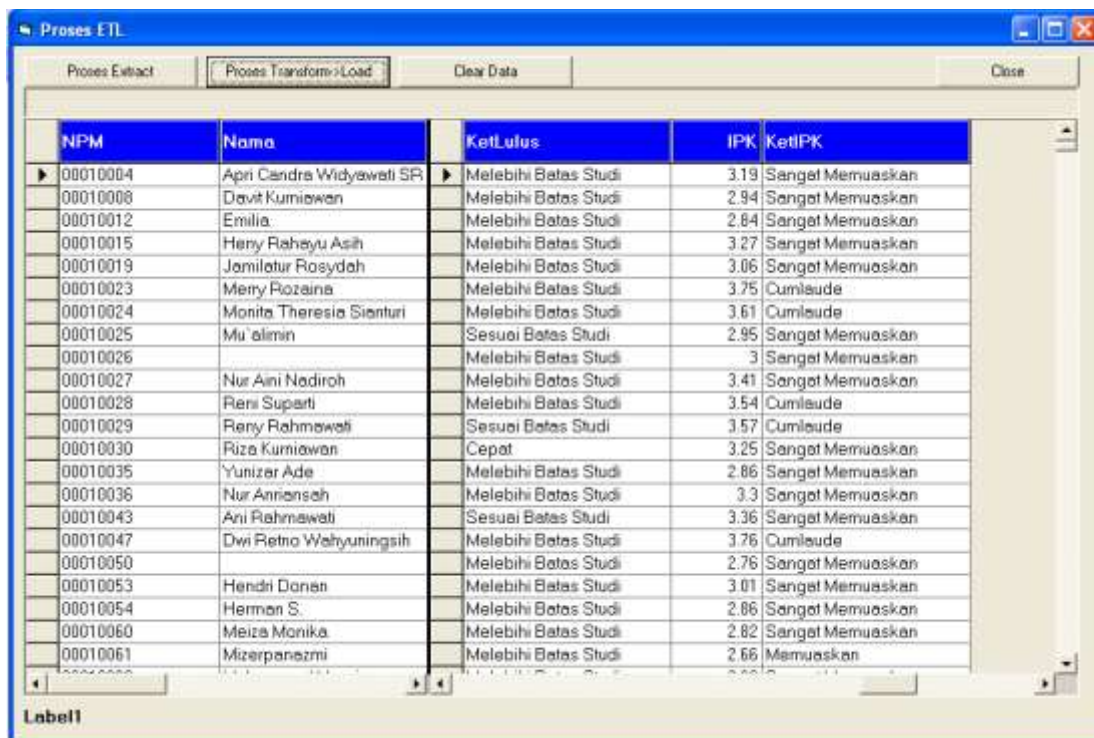




NPM	Nama	TahunLulus	LamaStudi	JurusanAlumni	KetLulus
00010004	Apri Candra Widyawati SR	2005	1	S1 Teknik Informatika	OK
00010008	David Kurniawan	2005	1	S1 Teknik Informatika	OK
00010012	Emilia	2005	1	S1 Teknik Informatika	OK
00010015	Hery Rahayu Asih	2005	1	S1 Teknik Informatika	OK
00010019	Jamilatur Rosyidah	2005	1	S1 Teknik Informatika	OK
00010023	Merry Rozaina	2005	1	S1 Teknik Informatika	OK
00010024	Monita Theresia Sianturi	2008	1	S1 Teknik Informatika	OK
00010025	Mu' alimin	2004	1	S1 Teknik Informatika	OK
00010026		2006	1	S1 Teknik Informatika	OK
00010027	Nur Aini Nadiroh	2005	1	S1 Teknik Informatika	OK
00010028	Reni Suparti	2005	1	S1 Teknik Informatika	OK
00010029	Reny Rahmawati	2004	1	S1 Teknik Informatika	OK
00010030	Riza Kurniawan	2003	1	S1 Teknik Informatika	OK
00010035	Yunizer Ade	2005	1	S1 Teknik Informatika	OK
00010036	Nur Annisah	2005	1	S1 Teknik Informatika	OK
00010043	Ani Rahmawati	2004	1	S1 Teknik Informatika	OK
00010047	Dwi Retno Wahyuningsih	2008	1	S1 Teknik Informatika	OK
00010050		2006	1	S1 Teknik Informatika	OK
00010053	Hendri Donen	2005	1	S1 Teknik Informatika	OK
00010054	Herman S.	2005	1	S1 Teknik Informatika	OK
00010060	Meiza Monika	2005	1	S1 Teknik Informatika	OK
00010061	Mizerpanozmi	2005	1	S1 Teknik Informatika	OK

Gambar 3. Hasil Proses *Extract*

Proses transformasi dapat dilakukan dengan cara mengklik tombol *Proses Transform → Load* seperti dapat dilihat pada gambar 4.



NPM	Nama	KetLulus	IPK	KetIPK
00010004	Apri Candra Widyawati SR	Melebihi Batas Studi	3.19	Sangat Memuaskan
00010008	David Kurniawan	Melebihi Batas Studi	2.94	Sangat Memuaskan
00010012	Emilia	Melebihi Batas Studi	2.84	Sangat Memuaskan
00010015	Hery Rahayu Asih	Melebihi Batas Studi	3.27	Sangat Memuaskan
00010019	Jamilatur Rosyidah	Melebihi Batas Studi	3.06	Sangat Memuaskan
00010023	Merry Rozaina	Melebihi Batas Studi	3.75	Cumlaude
00010024	Monita Theresia Sianturi	Melebihi Batas Studi	3.61	Cumlaude
00010025	Mu' alimin	Sesuai Batas Studi	2.95	Sangat Memuaskan
00010026		Melebihi Batas Studi	3	Sangat Memuaskan
00010027	Nur Aini Nadiroh	Melebihi Batas Studi	3.41	Sangat Memuaskan
00010028	Reni Suparti	Melebihi Batas Studi	3.54	Cumlaude
00010029	Reny Rahmawati	Sesuai Batas Studi	3.57	Cumlaude
00010030	Riza Kurniawan	Cepat	3.25	Sangat Memuaskan
00010035	Yunizer Ade	Melebihi Batas Studi	2.86	Sangat Memuaskan
00010036	Nur Annisah	Melebihi Batas Studi	3.3	Sangat Memuaskan
00010043	Ani Rahmawati	Sesuai Batas Studi	3.36	Sangat Memuaskan
00010047	Dwi Retno Wahyuningsih	Melebihi Batas Studi	3.76	Cumlaude
00010050		Melebihi Batas Studi	2.76	Sangat Memuaskan
00010053	Hendri Donen	Melebihi Batas Studi	3.01	Sangat Memuaskan
00010054	Herman S.	Melebihi Batas Studi	2.86	Sangat Memuaskan
00010060	Meiza Monika	Melebihi Batas Studi	2.82	Sangat Memuaskan
00010061	Mizerpanozmi	Melebihi Batas Studi	2.66	Memuaskan

Gambar 4. Hasil Transformasi

## PROSES MINING

Setelah data yang dihasilkan dari proses ETL selesai dilakukan, langkah selanjutnya adalah melakukan proses *mining*. Proses *mining* ini dilakukan dengan bantuan DBMS MS. SQL Server 2008 yang didalamnya telah terdapat fasilitas Data Mining. Fasilitas yang digunakan dalam SQL Server 2008 adalah *Analysis Services Toolset*. Fasilitas ini dikemas dalam Business Intelligence sebagai sebuah solutions dalam Microsoft Visual Studio 2008.

Untuk membuat sebuah project Data Mining ada beberapa tahapan yang perlu dilakukan, yaitu:

- 1) Membuat Koneksi Ke Sumber Data  
Koneksi digunakan untuk mengetahui sumber database yang ada di server. Informasi koneksi ini nantinya disimpan di dalam Data Source Designer.
- 2) Membuat Data Source View (DSV)  
Informasi dalam Data Source View (DSV) berisi tabel apa yang digunakan dalam proses mining.
- 3) Membuat Mining Structures  
Tahapan ini berisi informasi mengenai tabel yang digunakan dalam proses mining berikut kolom-kolom input dan kolom prediksi.

### 3.2 PEMBAHASAN

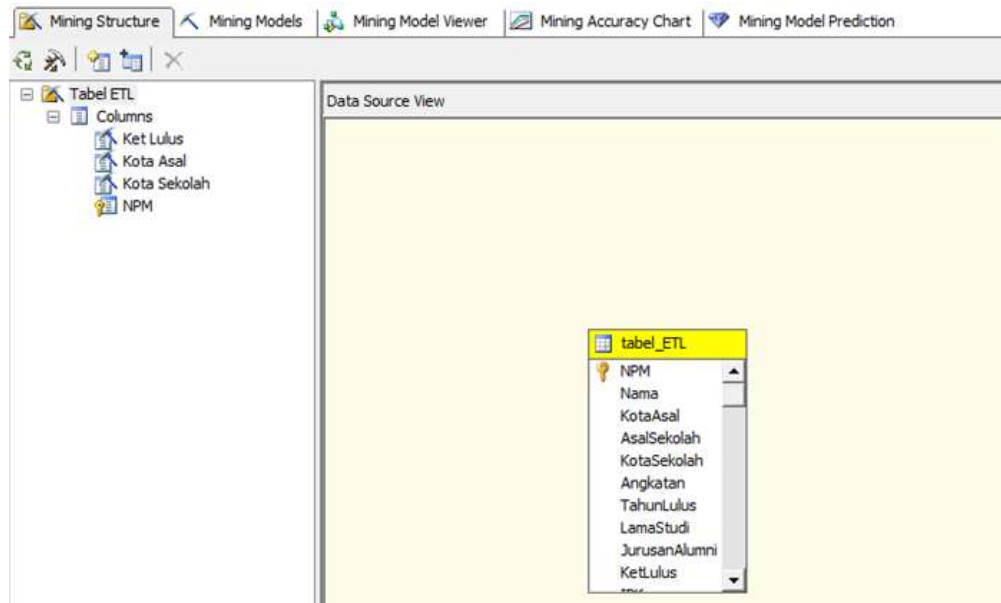
Proses *mining* yang dilakukan menggunakan sebuah model *mining structure* dengan Input berupa Kota asal dan asal sekolah, kemudian satu kolom predict yaitu KetLulus. Model mining ini digunakan untuk mengetahui hubungan (*asosiasi*) antara asal mahasiswa dan asal sekolah dengan tingkat kelulusan. Model aturan (*rule*) yang dibentuk nantinya adalah

if [asal sekolah] dan [kota asal] then [tingkat kelulusan]

Sebuah predikat yang digunakan dalam sebuah *rule* disebut sebagai *item*. Sebagai contoh dalam rule di atas asal sekolah merupakan sebuah item, demikian juga kota asal merupakan sebuah item. Item-item yang ada dalam rule selanjutnya disebut sebagai *itemset*. Sehingga dalam sebuah rule akan terdiri dari beberapa item yang membuat sebuah kondisi dan konklusi atau hubungan sebab-akibat.

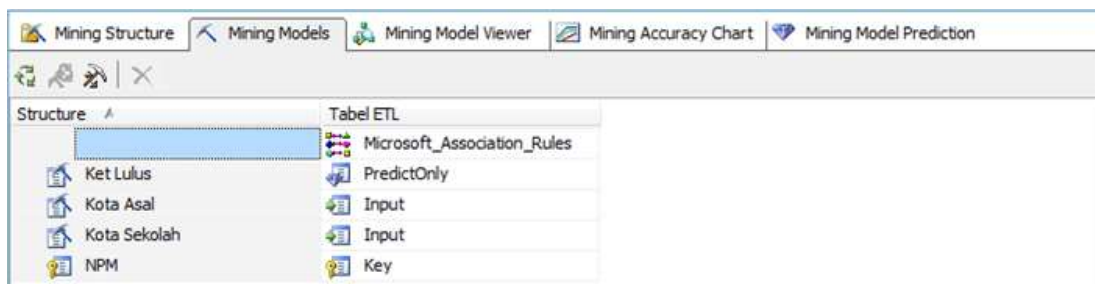
Algoritma *Association Rules* tidak menerima *attribute continuous* dikarenakan algoritma ini merupakan algoritma perhitungan yang menghubungkan antar attribute yang ada dalam sebuah rules. Di dalam association rules terdapat Dependency netview yang berguna untuk melakukan eksplorasi data yang melibatkan kondisi masing-masing attribute.

Model mining structure yang digunakan dalam penelitian ini dapat dilihat pada gambar 5.



Gambar 5. Model Mining Structure

Dari mining *structure* yang dibentuk, kemudian diperoleh mining model seperti dilihat pada gambar 6.

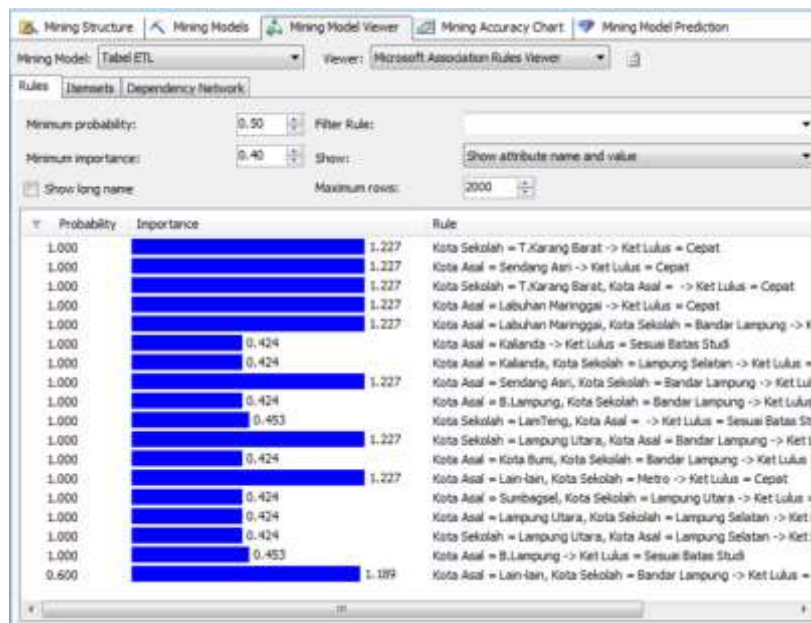


Gambar 6. Mining Models

Dalam proses mining terdapat sebuah kolom yang berfungsi sebagai *primary key*. Selanjutnya ditentukan kolom mana saja yang berfungsi sebagai *input* dan *predict*. Kolom predict ini berfungsi sebagai tujuan (*goal*) dari algoritma *asosiation rules*.

Untuk mendapatkan aturan-aturan dalam asosiation rules, selanjutnya data diproses dan dari menu *browser* dapat diperoleh gambaran dari hasil proses mining. Gambar 4.7 memperlihatkan sejumlah *rules* yang ada dari proses mining. *User* cukup memasukkan *minimum probability* dan *minimum importance*. Kemudian secara otomatis terbentuk sejumlah *rules* yang memenuhi nilai ambang batas (*threshold*) yang telah ditentukan.

Jika *minimum probability* (nilai *confidence*) diset dengan nilai yang kecil, maka jumlah *rules* yang dihasilkan akan semakin banyak. Hal ini berlaku juga untuk *minimum importance*. Importance support digunakan untuk mengetahui seberapa pentingkah sebuah rule digunakan, semakin tinggi nilai *importance* maka *rule* tersebut semakin dapat digunakan. Importance atau biasa disebut dengan interesting *score* atau *lift* dalam beberapa *literature* dapat digunakan untuk mengukur jumlah itemset dan *rules*.



Gambar 7. Rules Viewer

Sebagai contoh terbentuk rules :

Kota sekolah=T. Karang Barat→ Ket Lulus=Cepat

Aturan ini memberi arti bahwa kemungkinan mahasiswa yang kota sekolahnya berasal dari Tanjung Karang Barat akan lulus cepat. Hal ini diketahui dengan kemungkinan atau tingkat *confidence* yang diwakili oleh nilai *Probability* 1. Hanya aturan-aturan yang nilai *probability* di atas 0.50 (sesuai dengan nilai *probability* yang ditentukan) yang akan ditampilkan. Demikian juga dengan nilai *Minimum Importance*, hanya *rules* yang memiliki nilai di atas nilai minimum importance yang ditampilkan. Aturan-aturan yang muncul tersebut dapat digunakan untuk memprediksi tingkat kelulusan mahasiswa.

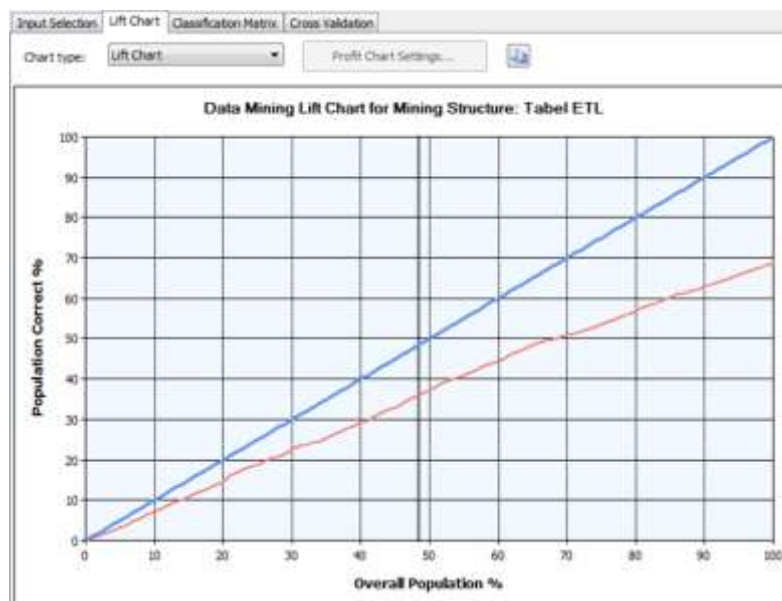
Selanjutnya untuk mengetahui seberapa besar tingkat dukungan item atau kolom-kolom yang ada dalam sebuah frequent item, dapat digunakan tab Itemset seperti dapat dilihat pada gambar 8. Dalam gambar ini setiap *itemset* memiliki *size* atau ukuran. *Size* mengindikasikan seberapa banyak item yang ada dalam sebuah itemset. Nilai *minimum support* digunakan untuk membatasi itemset bukan untuk membatasi jumlah *rules*. Sedangkan *minimum probability* tidak berdampak pada itemset melainkan berdampak pada jumlah *rules*.





Node tujuan (*goal*) akan berisi sejumlah nilai pada atribut tujuan yang telah ditentukan. Missal dalam atribut *predict* ada 1 kolom dan berisi 3 buah data, maka jumlah *node goal* akan berisi 3 buah. Masing-masing node akan terhubung dengan atribut yang ditentukan sebagai kolom input pada *mining structure*. Masing-masing jalur mewakili pasangan asosiasi dari sebuah aturan. Semakin rendah nilai *minimum probability* maka jumlah node dan jalur akan semakin banyak.

Untuk mengetahui seberapa besar tingkat akurasi data jika aturan-aturan yang telah terbentuk digunakan untuk melakukan prediksi dan mewakili seluruh populasi dapat dilihat pada tab *lift chart* seperti dilihat pada gambar 10.



Gambar 10. *Lift Chart*

*Rule* yang terbentuk dari item yang ada mempunyai nilai *confidence* yang berbeda. *Confidence* merupakan tingkat kepercayaan atau tingkat kebenaran dari *rule* yang terbentuk. *Rule* yang dicari adalah yang mempunyai kemungkinan benar yang besar, yaitu *rule* yang membuat prediksi yang benar (atau sangat sering benar).

Dari tabel diatas dapat diketahui, bahwa jika dengan menggunakan nilai *maximum support* lebih besar, maka nilai *confidence* yang diperoleh juga lebih besar.

## 4. KESIMPULAN DAN SARAN

### 4.1 KESIMPULAN

Kesimpulan yang dapat diambil adalah Aplikasi Data Mining ini dapat digunakan untuk menampilkan informasi tingkat kelulusan. Informasi yang ditampilkan berupa nilai *support* dan *confidence* hubungan antara tingkat kelulusan dengan data induk mahasiswa. Semakin tinggi nilai *confidence* dan *support* maka semakin kuat nilai hubungan antar atribut. Data induk mahasiswa yang diproses mining meliputi data asal sekolah, data kota mahasiswa. Hasil dari proses data mining ini dapat digunakan sebagai pertimbangan dalam

mengambil keputusan lebih lanjut tentang faktor yang mempengaruhi tingkat kelulusan khususnya faktor dalam data induk mahasiswa.

## 4.2 SARAN

Dari penelitian yang telah peneliti lakukan, masih perlu diadakan penelitian lebih lanjut tentang data-data yang digunakan dalam prediksi tingkat kelulusan. Sehingga parameter yang digunakan lebih mengena, jadi data yang dihasilkan lebih dapat digunakan sebagai alat pendukung bagi pihak yang bersangkutan untuk mengambil kebijakan dan sebagai sumber informasi yang lebih akurat untuk menemukan ide-ide baru yang lebih baik.

## REFERENSI

- [1] Bramer, Max, 2007, "*Principles of Data Mining*", Springer, London.
- [2] Chintakayala, Padmini. 2005. "*Beginners Guide for Software Testing : Symbiosys Technologies*".
- [3] David Hand, Heikki Mannila, and Padhraic Smyth, "*Principles of Data Mining*", MIT Press, Cambridge, MA, 2001.
- [4] Han, J. and Kamber, M, 2006, "*Data Mining Concepts and Techniques Second Edition*". Morgan Kauffman, San Francisco. s
- [5] Kusriani, dan Emha Taufik Luthfi, 2009, "*Algoritma Data Mining*", Penerbit Andi, Yogyakarta.
- [6] Santosa, Budi, 2007, "*Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis*", Graha Ilmu, Yogyakarta.
- [7] Sommerville, Ian, 2003, "*Software Engineering (Rekayasa Perangkat Lunak)/ Edisi 6/Jilid 1*" Erlangga, Jakarta.
- [8] Witten, I. H and Frank, E. 2005. "*Data Mining : Practical Machine Learning Tools and Techniques Second Edition*". Morgan Kauffman : San Francisco.
- [9] The Gartner Group, URL:<http://www.gartner.com>